What is Claimed is:

[c1] A light weight subject indexing system, comprising:

a candidate headword identification system for identifying candidate words in the subject line of a document which are not listed in a user modified common word list;

a lexical context system for creating a lexical context for an identified candidate headword;

a ranking system for ranking the set of identified candidate headwords for a collection of documents and selecting among them for inclusion in an index; and

an index creation system for listing selected candidate headwords based on the results of ranking and selection.

[c2] The system of claim 1, further comprising:

a user modified common word list system for creating the user modified common word list by modifying a list of common words by at least one of adding words to the common word list, subtracting words from the common word list, or both adding and subtracting words from the common word list.

[c3] The system of claim 1, wherein the candidate headword identification system scans the subject line of a document to identify as a candidate headword each word in the subject line that does not appear in the user modified common word list, and, for each candidate headword found in the subject line, adds the candidate headword to the accumulated list of candidate headwords for the collection if the word has not been previously identified as the candidate headword for the collection, associates the entry for the word in the accumulated list with the subject line, and increments subject line occurrence counts for the word.

[c4] The system of claim 1, wherein the lexical context system identifies the lexical context for the candidate headword as the words to the left and the right of the candidate headword up to, but not including, a barrier word.

[c5] The system of claim 4, wherein if there are no non-barrier words to the

immediate left and the immediate right of the candidate headword, but there are non-barrier words beyond the barrier words either to the left or to the right of the candidate headword, the lexical context system identifies, as the lexical context for the headword, the barrier words in the direction in which the non-barrier words appear, plus the non-barrier words beyond those barrier words up to, but not including, the next barrier word.

[c6] The system of claim 4 , wherein if there are no non-barrier words to the immediate left and the immediate right of the candidate headword, but there are non-barrier words beyond the barrier words both to the left and to the right of the candidate headword, the lexical context system uses heuristic means to determine which direction is to be used in establishing the lexical context, and identifies the lexical context as consisting of the barrier words immediately following the candidate headword in that direction, plus the words beyond those barrier words up to, but not including, the next barrier word.

[c7] The system of claim 4 , wherein if the subject line contains only barrier words in addition to the candidate headword, the lexical context system may use the barrier words as lexical context.

[c8] The system of claim 1, wherein the ranking system ranks the candidate headwords based on count information obtained during candidate headword selection.

[c9] The system of claim 1, wherein the ranking system ranks the candidate headwords based on the number of unique subject lines in which a candidate headword occurs and the number of individual messages in which the candidate headword occurs in a subject line.

[c10] The system of claim 1, wherein the ranking system selects the highest ranking headwords up to a desired index size.

[c11] The system of claim 1, wherein the index creation system lists each selected candidate headword in a predetermined order followed by the lexical contexts in which the candidate headword appears.

[c12] The system of claim 1, wherein the index creation system lists each selected candidate headword in a predetermined order followed by the subject lines in which the candidate headword appears.

[c13] The system of claim 1, wherein the index creation system links the candidate headword to a representation of the set of messages in whose subject lines the candidate headword appears.

[c14] The system of claim 1, wherein the index creation system links the lexical contexts in which the candidate headword appears to the set of messages in whose subject lines the lexical contexts appear.

[c15] The system of claim 1, wherein the index creation system lists the selected candidate headwords in a user specified order.

[c16] The system of claim 1, wherein the index creation system limits the number of lexical contexts that are listed below the candidate headword.

[c17] A method for creating a light weight subject index, comprising:

identifying, as candidate headwords, words in the subject lines of a collection of documents which are not listed in a user modified common word list;

creating lexical contexts for identified candidate headwords;

ranking the set of identified candidate headwords for a collection of documents and selecting among them for inclusion in an index; and

listing selected candidate headwords based on the results of ranking and selection.

[c18] The method of claim 17, further comprising:

creating the user modified common word list by modifying a list of common words by adding words to and/or subtracting words from the common word list.

[c19] The method of claim 17, wherein the candidate headwords for a document are identified by:

scanning the subject line of the document to identify, as candidate

headwords, those words that do not appear in the user modified common word list, and for each such word found;

adding the word to an accumulated list of candidate headwords for the collection if the word has not been previously identified as a candidate headword;

associating the entry for the word in the accumulated list with the subject line; and

incrementing subject line occurrence counts for the word.

[c20] The method of claim 17, wherein the lexical context for a candidate headword within a subject line is identified as the words to the left and the right of the candidate headword up to, but not including, a barrier word.

[c21] The method of claim 20 wherein if there are no non-barrier words to the immediate left and immediate right of the candidate headword, but there are non-barrier words beyond one or more barrier words either to the left or to the right of the candidate headword, the lexical context for the candidate headword is identified as the barrier words in the direction in which the non barrier words appear, plus the non-barrier words beyond those barrier words up to, but not including, the next barrier word.

[c22] The method of claim 20, wherein if there are no non-barrier words to the immediate left and immediate right of the candidate headword, but there are non-barrier words beyond the barrier words both to the left and to the right of the candidate headword, heuristic means are used to determine which direction is to be used in establishing the lexical context, and the lexical context is identified as consisting of the barrier words immediately following the candidate headword in that direction, plus the words beyond those barrier words up to, but not including, the next barrier word.

[c23] The method of claim 20, wherein if no content words are found on both the left and right of the candidate headword, the lexical context may be identified as including those barrier words.

[c24] The method of claim 17, wherein the candidate headwords are ranked based on

count information obtained during candidate headword selection.

[c25] The method of claim 17, wherein the candidate headwords are ranked based on the number of unique subject lines in which a candidate headword occurs and the number of individual messages in which the candidate headword occurs in a subject line.

[c26] The method of claim 17, wherein the highest ranking headwords are selected up to a desired index size.

[c27] The method of claim 17, wherein each selected candidate headword is listed in a predetermined order followed by the lexical contexts in which the candidate headword appears.

[c28] The method of claim 17, wherein each selected candidate headword is listed in a predetermined order followed by the subject lines in which the candidate headword appears.

[c29] The method of claim 17, wherein the candidate headword is linked to a representation of the set of messages in whose subject lines the candidate headword appears.

[c30] The method of claim 17, wherein the lexical context in which a candidate headword appears is linked to a representation of the set of messages in whose subject lines the lexical context appears.

[c31] The method of claim 17, wherein the candidate headwords are listed in a user specified order.

[c32] The method of claim 17, wherein the number of lexical contexts that are listed below the candidate headword is limited.

[c33] A system for creating a user specified index, comprising:

at least one user interface for specifying a desired index;

a document application system electrically connected to the at least one user interface; and

an indexing system for creating the desired index, the indexing system

comprising:

a candidate headword identification system for identifying candidate words in the subject line of a document which are not listed in a user modified common word list;

a lexical context system for creating a lexical context for an identified candidate headword;

a ranking system for ranking the set of identified candidate headwords for a collection of documents and selecting among them for inclusion in an index; and

an index creation system for listing selected candidate headwords based on the results of ranking and selection.

[c34] The system of claim 33, wherein the documents are stored in an archive.

[c35] The system of claim 33, wherein the document application system is a list server or a personal email application directly associated with and controlled by one of the at least one user interface.

[c36] The system of claim 33, wherein the indexing system is integrated with the document application system.

[c37] The system of claim 33, wherein the at least one user interface, the document application system and the indexing system are electrically connected by at least one communication link.